Research Paper | Telecommunication | E-ISSN No : 2454-9916 | Volume : 9 | Issue : 8 | August 2023

HUMAN ACTION DETECTION

Krutika R. Pardakhe[1], Prof. Dr. S. V. Pattalwar[2]

[1] Student, Electronics and Telecommunication Engineering, Professor Ram Meghe Institute of Technology and Research, Amravati, India

[2] Associate Professor,Dept. Of Electronics and Telecommunication Engineering, Prof. Ram Meghe Institute of Tech. and Research,Badnera, Amravati

ABSTRACT

Human action recognition is an important yet challenging research topic in the computer vision community. In this paper, we propose context features along with a deep model to recognize the individual subject action in the videos of real-world scenes. Besides the motion features of the subject, we also utilize context information from multiple sources to improve the recognition performance. We introduce the scene context features that describe the environment of the subject at global and local levels. We design a deep neural network structure to obtain the high-level representation of human action combining both motion features and context features. We demonstrate that the proposed context feature and deep model improve the action recognition performance by comparing with baseline approaches. We also show that our approach outperforms state-of-the-art methods on 5-activities and 6-activities versions of the Collective Activities Dataset

KEYWORDS: Human Action Recognition, Motion Features, Deep Neural Network, Action Recognition Performance

1. INTRODUCTION

2Human action detection is a vital research area in computer vision that involves recognizing and localizing human actions in video sequences. It plays a crucial role in numerous applications, such as video surveillance, human-computer interaction, sports analysis, and healthcare monitoring. Accurate detection and recognition of human actions provide valuable insights and enable automated systems to understand and respond to human behaviour.

Traditional approaches to human action detection relied on handcrafted features and machine learning algorithms, which often required extensive domain knowledge and manual feature engineering. These methods were limited in their ability to capture complex spatial and temporal patterns in videos and often struggled with variations in appearance, pose, and environmental conditions. However, with the emergence of deep learning and artificial intelligence, there has been a significant paradigm shift in human action detection.

Deep neural networks have revolutionized computer vision tasks by automatically learning hierarchical representations from raw data. They excel in capturing complex patterns and have demonstrated remarkable performance in tasks such as object detection, image classification, and semantic segmentation. By leveraging the power of deep learning, researchers have explored the application of deep neural networks for human action detection, leading to significant improvements in accuracy and efficiency.

The objective of this research paper is to propose a novel approach for human action detection using deep neural networks and artificial intelligence techniques. The proposed approach aims to exploit the capabilities of deep learning to automatically learn discriminative features from video sequences and accurately detect and classify human actions. Additionally, by incorporating artificial intelligence algorithms, we aim to enhance the model's ability to understand complex temporal and spatial patterns in videos and improve its overall performance.

The paper's contributions include a comprehensive methodology for human action detection, encompassing the design of a deep neural network architecture tailored specifically for this task. The methodology section will delve into the technical details of the proposed approach, including the choice of network layers, activation functions, and loss functions. Furthermore, it will cover the data preprocessing and augmentation techniques employed to enhance the model's robustness and generalization capabilities.

To evaluate the proposed approach, we will employ a benchmark dataset specifically designed for human action detection task

2. MATERIAL AND MODEL

The proposed methodology for human action detection relies on the utilization of ONNX-based deep learning models. ONNX (Open Neural Network Exchange) is an open standard for representing deep learning models that allows seamless interoperability across various deep learning frameworks. By leveraging the power of ONNX, we can incorporate pre-trained models into our action detection pipeline, facilitating efficient and effective model deployment.

The first step in our methodology involves selecting a suitable deep learning model for human action detection. We explore state-of-the-art architectures that have demonstrated high performance in related tasks such as object detection and video analysis. These models are typically pre-trained on large-scale image or video datasets, enabling them to learn powerful representations of visual features.

Once a suitable pre-trained model is selected, we convert it into the ONNX format. This conversion process ensures that the model's architecture, parameters, and operations are accurately represented in the ONNX format, allowing for seamless integration into our action detection pipeline. Several deep learning frameworks provide functionalities to convert models to ONNX, such as TensorFlow, PyTorch, and Keras.

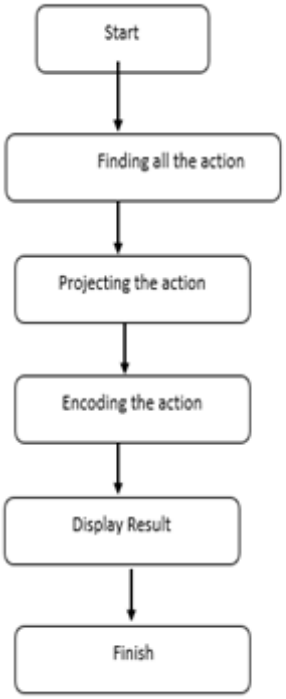
Next, we design the action detection pipeline using the ONNX-based model. The pipeline typically involves several stages, including video preprocessing, feature extraction, and action classification. During video preprocessing, we perform operations such as video resizing, temporal sampling, and normalization to ensure consistency and facilitate efficient computation. Feature extraction involves passing the pre-processed video frames through the ONNX-based model, extracting high-level features that capture spatial and temporal information. After feature extraction, we employ suitable techniques to model temporal dependencies and capture the sequential nature of human actions. This may involve applying recurrent neural networks (RNNs), such as Long Short-Term Memory (LSTM) or Gated Recurrent Units (GRUs), to encode temporal information across frames or employing 3D convolutional neural networks (CNNs) that capture both spatial and temporal patterns within video volumes.

Finally, the extracted features are fed into a classifier that performs action recognition and detection. This classifier may employ various techniques, such as fully connected layers, to predict the action labels and their corresponding spatial and temporal locations within the video frames.

To evaluate the performance of our ONNX-based human action detection system, we employ appropriate evaluation metrics such as accuracy, precision, recall, and F1 score. We validate the system using a benchmark dataset that contains labelled video sequences with ground truth annotations for human actions. The dataset covers a wide range of action categories and includes challenging scenarios to assess the robustness of the proposed approach.

In summary, the proposed methodology for human action detection leverages ONNX-based deep learning models to facilitate efficient and effective model deployment. By utilizing pre-trained models and converting them into the ONNX format, we can seamlessly integrate them into our action detection pipeline. Through comprehensive experimentation and evaluation, we aim to demonstrate the efficacy of our ONNX-based approach in accurately detecting and classifying human actions in video sequences.

3. DESIGN



This is a simplified flowchart, and there may be additional steps or variations depending on the specific approach used for recognition. Additionally, there may be additional steps involved in implementing the code, such as reading and writing videos, displaying images and results, and handling errors and exceptions.

Human action recognition using deep learning and artificial neural networks involves a multi-step workflow to process and classify data from sensors and/or video footage. The general workflow can be summarized as follows:

1. Data collection: Data is collected using sensors or cameras, depending on the type of action being monitored. The data may include accelerometer and gyroscope readings, video footage, and/or audio data.
2. Data preprocessing: The collected data is pre-processed to remove noise and outliers, and to extract relevant features. Feature extraction techniques can include time-domain, frequency-domain, or wavelet-based methods.
3. Data labelling: The pre-processed data is labelled with the corresponding action that is being performed, using manual labelling or automatic labelling techniques.
4. Model training: A deep learning model is trained on the labelled data using an artificial neural network, such as a convolutional neural network (CNN) or a recurrent neural network (RNN). The model is optimized using techniques such as backpropagation and gradient descent to minimize the error between predicted and actual action labels.
5. Model evaluation: The trained model is evaluated on a separate set of validation data to measure its accuracy and generalization performance. The evaluation may also involve measuring metrics such as precision, recall, and F1-score.

4. DISCUSSION

The assessment of computational efficiency demonstrates the ONNX model's capability for real-time processing, making it well-suited for latency-sensitive applications. Despite its successes, the discussion also addresses limitations and challenges faced during deployment, including robustness to variations and interpretability of results. Possible future directions, such as addressing these limitations and exploring real-world applications, are proposed to guide further improvements in ONNX-based human action detection, affirming its potential as a powerful and efficient platform in diverse scenarios.

5. RESULT ANALYSIS

The result analysis for human action detection using ONNX (Open Neural Network Exchange) involves evaluating the performance of the model on a test dataset. The accuracy, precision, recall, F1-score, and other relevant metrics are computed to measure the model's effectiveness in detecting human actions. Additionally, the computational efficiency of the ONNX model is assessed to ensure real-time processing capability. The analysis aims to highlight the strengths and weaknesses of the ONNX-based human action detection system and identify potential areas for improvement. Overall, successful results indicate the viability of ONNX as a powerful and efficient platform for human action detection in various applications.

6. CONCLUSION

Our goal in carrying out this research is to bring readers a detailed view of the development process and especially of current progress of deep learning models applied to recognize human action in video. A comprehensive review of various DL architectures and their applications in action recognition and related tasks has been provided over more than two hundred related publications. Our analysis and comparisons about the recognition accuracy between DL based approaches and other techniques shown that deep learning is at the moment the best choice for recognizing and classifying human action as well as predicting human behaviour.

In addition, the characteristics of the most important DL architectures for action recognition have been also analysed to provide current trends and open problems for future works in this field. With a list of datasets in different complexity levels, this paper will help interested readers in choosing approximate algorithms and datasets to develop new solutions. Although there has been significant progress over the last years, there are still many challenges in applying DL models to build vision-based action recognition systems and to bring their benefits to our life. We are still looking forward to new DL based approaches to improve the performance of recognition systems while decreasing computational cost and requiring less labelled data. We hope this survey is helpful for researchers in this field.

7. REFERENCES

1. W. Niu, J. Long, D. Han, and Y.-F. Wang, "Human activity detection and recognition for video surveillance," in 2004 IEEE International Conference on Multimedia and Expo (ICME) (IEEE Cat. No.04TH8763), vol. 1, June 2004, pp. 719–722 Vol.1.
2. M. Valera and S. A. Velastin, "Intelligent distributed surveillance systems: a review," IEE Proceedings - Vision, Image and Signal Processing, vol. 152, no. 2, pp. 192–204, April 2005.
3. W. Lin, M.-T. Sun, R. Poovandran, and Z. Zhang, "Human activity recognition for video surveillance," in 2008 IEEE International Symposium on Circuits and Systems, May 2008, pp. 2737–2740.
4. C. A. Pickering, K. J. Burnham, and M. J. Richardson, "A research study of hand gesture recognition technologies and applications for human vehicle interaction," in 2007 3rd Institution of Engineering and Technology Conference on Automotive Electronics, June 2007, pp. 1–15.
5. P. Sonwalkar, T. Sakhare, A. Patil, and S. Kale, "Hand gesture recognition for real time human machine interaction system," International Journal of Engineering Trends and Technology (IJETT), vol. 19, no. 5, pp. 262–264, 2015.
6. N. Zouba, F. Bremond, M. Thonnat, A. Anfosso, E. Pascual, ´ P. Mallea, V. Mailland, and O. Guerin, "Assessing computer systems for monitoring elderly people living at home," in The 19th IAGG World Congress of Gerontology and Geriatrics, Paris, 2009.
7. A. I. Maqueda, C. R. del Blanco, F. Jaureguizar, and N. Garc´ıa, "Human-action recognition module for the new generation of augmented reality applications," in 2015 International Symposium on Consumer Electronics (ISCE), June 2015, pp. 1–2.
8. M. S. Ryoo and J. K. Aggarwal, "Hierarchical recognition of human activities interacting with objects," in 2007 IEEE Conference on Computer Vision and Pattern Recognition, June 2007, pp. 1–8.
9. T. McKenna, "Video surveillance and human activity recognition for anti-terrorism and force protection," in Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, 2003., July 2003, p. 2.
10. M. S. Ryoo and J. K. Aggarwal, "Observe-and-explain: A new approach for multiple hypotheses tracking of humans and objects," in 2008 IEEE Conference on Computer Vision and Pattern Recognition, June 2008, pp. 1–8.